



wellcome
connecting
science

computational course
**Computational systems biology for
complex human disease:
from static to dynamic
representations of disease mechanisms**

4–9 December 2022
Wellcome Genome Campus, UK

Introduction to Boolean modelling for biological systems

Dr Anna Niarakis

Hinxton Campus, Tuesday 6th December 2022

Basic concepts and theory

- Computational systems biology methods have widely employed dynamic models to describe the biological functions from the dynamic system point of view.
- **Keywords:** system / dynamic



What is the advantage of adding a dynamic layer?

provide
quantitative/qualitative
descriptions of the
network

predict the behavior of the
network under different
conditions, i.e., gene
knockout, treatment with
an external agent, etc.

Biological functions: Dynamic!

A biological function must be considered as a dynamic process

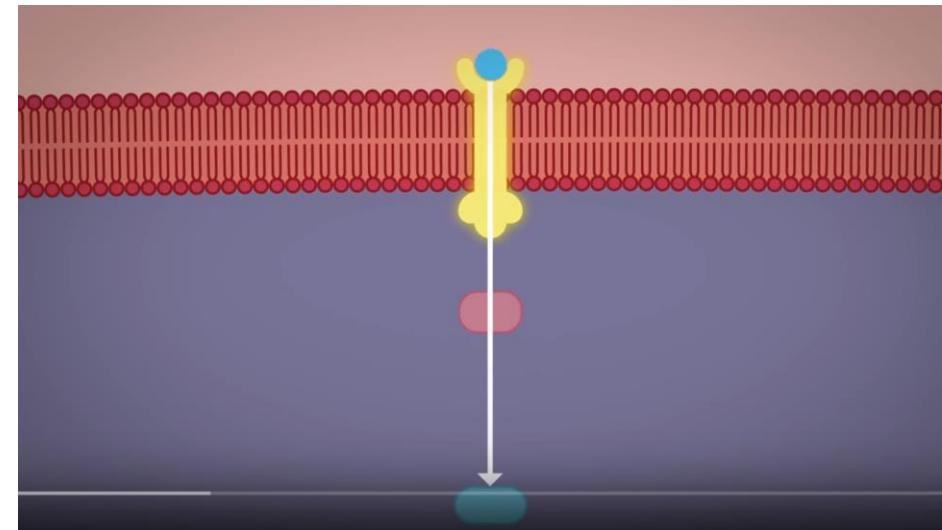
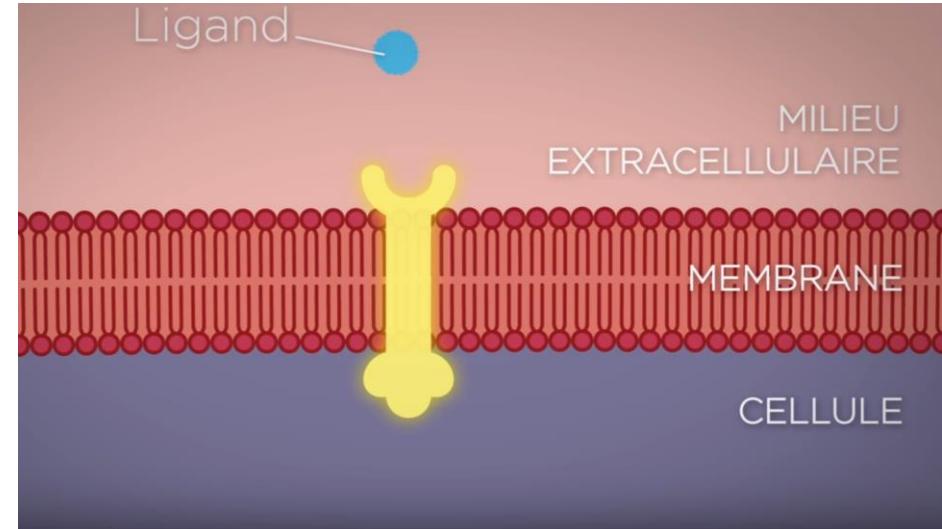
Within a cell, all the elements vary over time (energy production, respiration, formation of complexes, etc.)

For our explorations: A cell frozen in time t

After or before: Definitely different function!

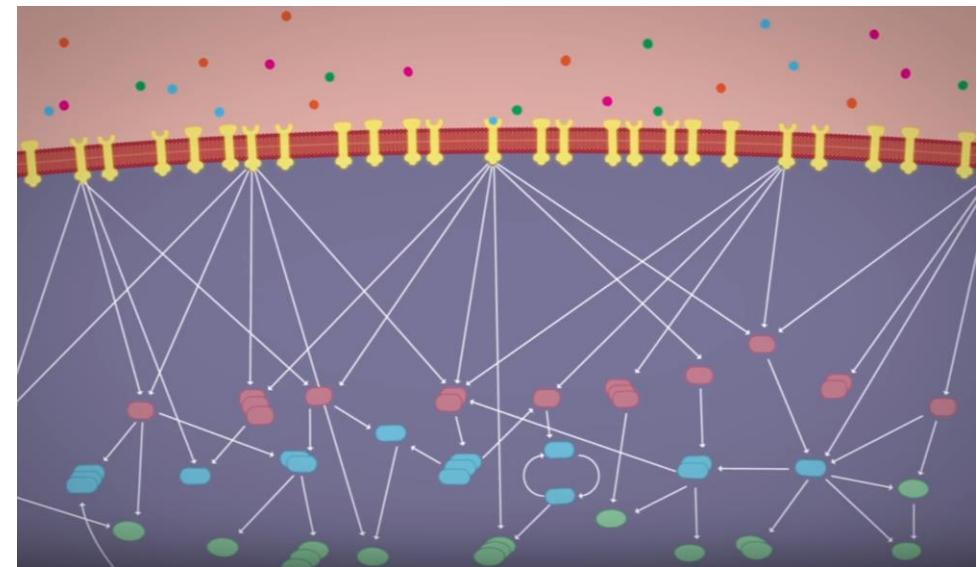
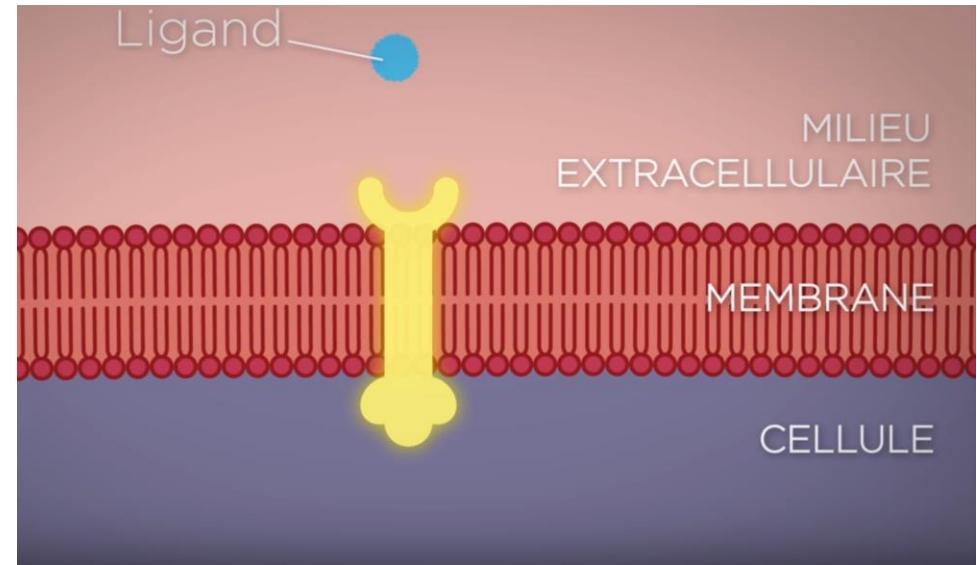
Many different layers

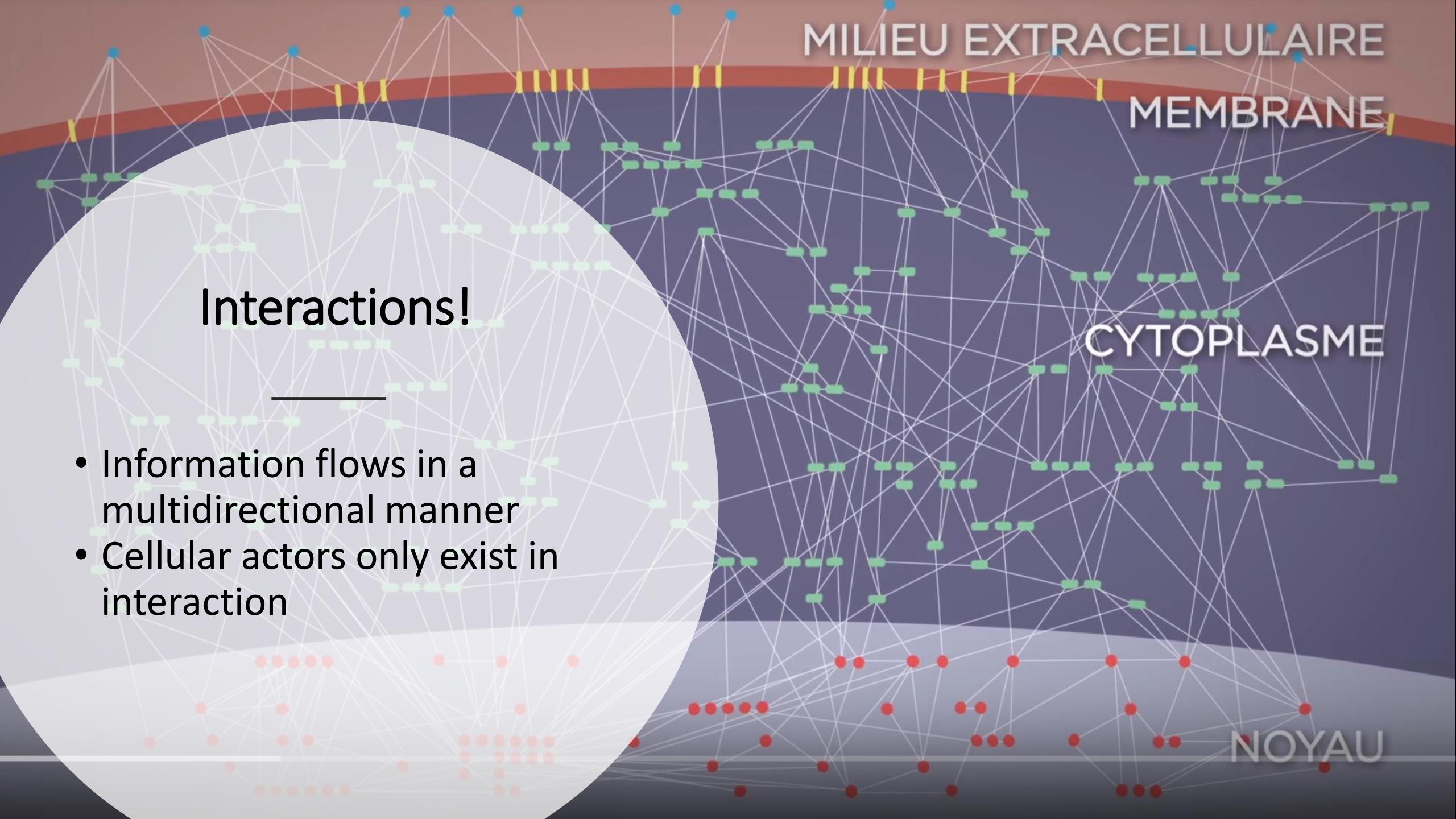
- Cellular functions involve multiple layers/ levels in the cell
- Functions cannot be defined independently of the context in which they operate



Several causes and consequences (effects)

- Several causes and several consequences coexist at the same time within the cell
- In the context of cellular interactions, it is impossible to associate a single cause with each consequence and a single consequence with each cause

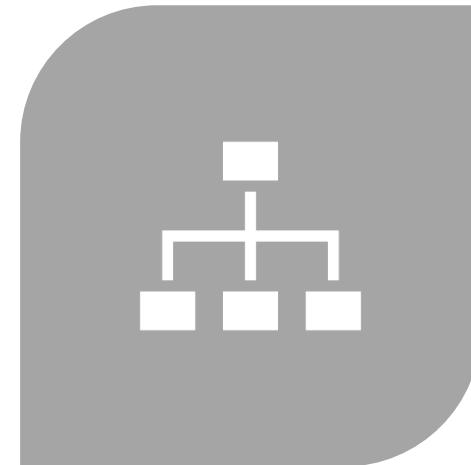




Cellular context: complex!



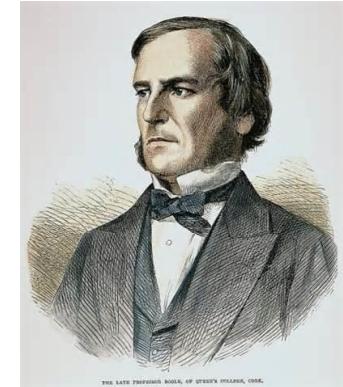
COMPLEXITY OF CELLULAR
CONTEXT:



ADOPT A DYNAMIC, SYSTEMS
POINT OF VIEW TO ADDRESS IT

Boolean Algebra

- Boolean Algebra is a mathematical system for formulating logical statements with appropriate symbols, so that logical problems can be solved algebraically (in a manner similar to ordinary algebra).
- Boolean Algebra is introduced by an English mathematician, George Boole (1815-1864).
- The widespread use of Boolean Algebra is initiated by an American mathematician, Claude Shannon (1916-2001).
- He is credited with founding both digital computer and digital circuit design theory.



• George Boole



• Claude Shannon

Boolean Algebra

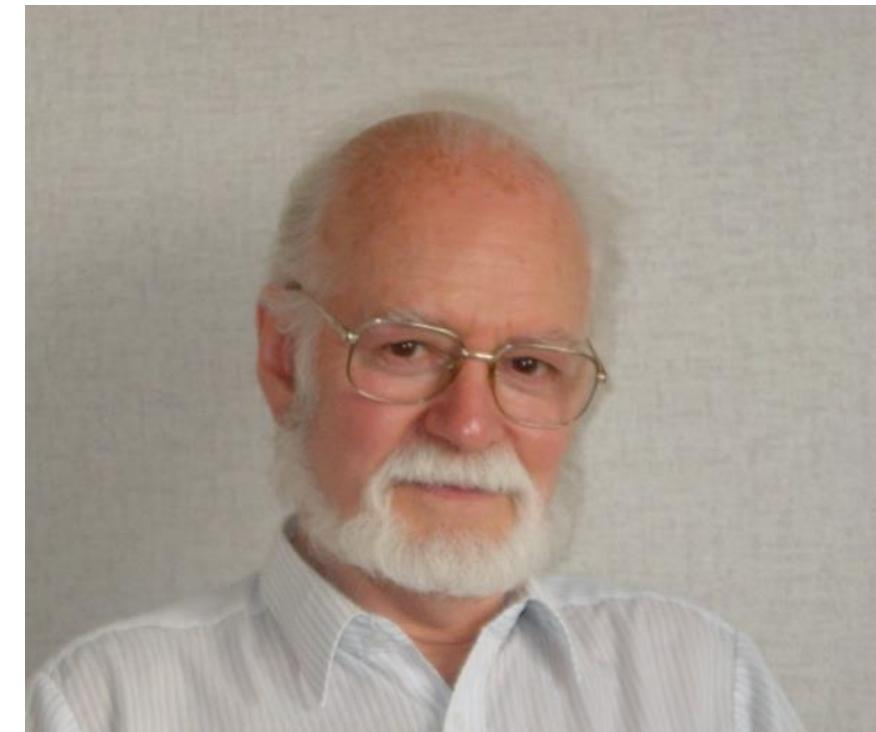
- The variables in Boolean Algebra are the truth values of True (1) and False (0).
- The main operations of Boolean Algebra are :
 - The conjunction (**and**)
 - The disjunction (**or**)
 - The negation (**not**)

The blackboard contains several mathematical and logical expressions:

- $D(x) = -3 + 3 + 4.31447$
- $\sqrt{a^2 + b^2} = x^2 + y^2$
- $x^2 + y^2 = ab + bc$
- $c(x, y) \left\{ \begin{array}{l} xy = c \\ cx - cy = 0 \\ \pi = c \end{array} \right.$
- $A \wedge B, A \Rightarrow B$
- $\frac{2x+3}{y} + \frac{x^2+3^2}{c} + \frac{2}{x} \rightarrow S$
- $\text{men} = 384 + n^{3v}$
- $x = 9.25$
- $\sum N_{30} \cdot x - \frac{1}{2} [964 + x]$
- $\beta = 9 + 3$

Boolean algebra for modelling regulatory networks

The experimental advances in the large-scale mapping of regulatory networks are fairly recent, but modeling efforts date back to the end of 1960s. Pioneering work of Stuart Kauffman and Rene Thomas.



In the absence of experimental results, Stuart Kauffman considered an idealized representation of a typical gene network

He assumed that genes are equivalent, and their interactions form a directed graph in which each gene receives inputs from a fixed number K of randomly selected neighbors.



Focus on asymptotic behaviour

Two types of attractors: stable states or simple cycles

Deterministic behaviour (only one possible following state)
=> Synchronous strategy

How does this work?

- The state of genes is described by binary (ON/OFF : 1/0) variables at a time point (t), and the dynamic behavior of each variable, that is, whether it will be ON (=1) or OFF (=0) at next moment ($t+1$), is governed by a Boolean function.
- A Boolean or logical function is written using the logical operators “and”, “or” and “not” to describe regulation.

Logical description of transcriptional regulation

René Thomas developed a detailed logical description of the mechanisms governing transcriptional regulation, including the effects of DNA domains such as promoters, initiators, terminators, and the concepts of genetic dominance and recessivity



Regulatory circuits: positive and negative

Non-deterministic behaviour
(more than one possibility following state)
=> Asynchronous strategy

Feedback loops and multivariate variables

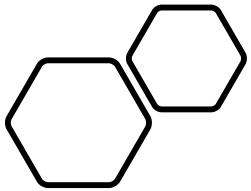
This formalism was later refined to include multilevel variables and used to study feedback loops, i.e. circular chains of interaction.



These loops can be classified into two categories based on the number of negative (inhibitory) interactions in the loop:

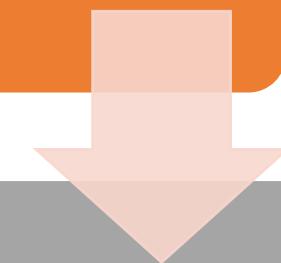
if this number is even,
the loop is positive,
and if the number of

negative interactions is
odd, the loop is a
negative feedback loop.



Associating feedback loops with dynamic behaviour and biological functions

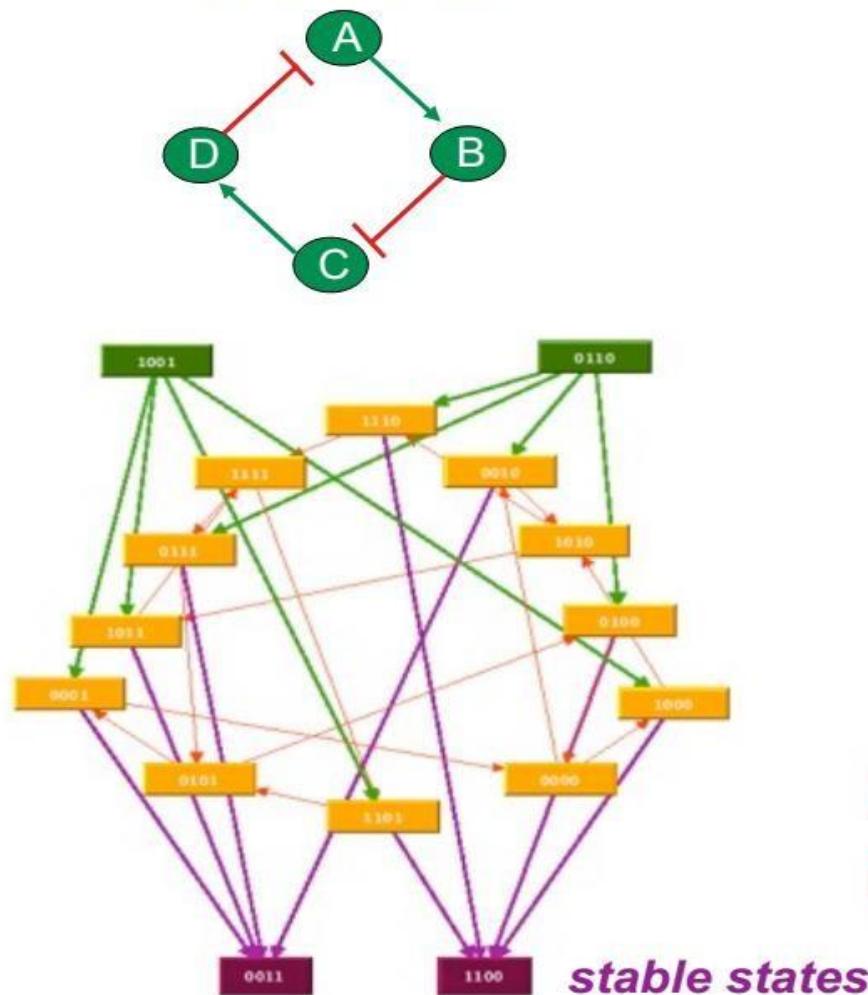
Thomas found that a positive feedback loop is a necessary condition for the existence of multiple steady states, while a negative feedback loop with two or more elements is a necessary condition for stable limit cycles .



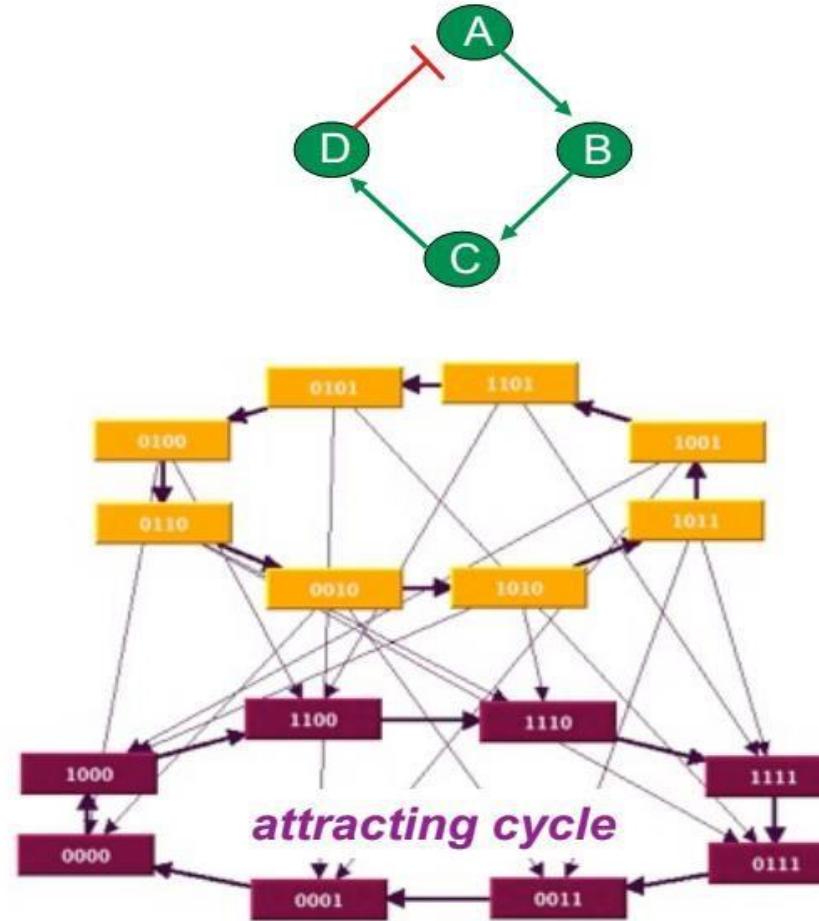
Biologically this means that cell differentiation is based on positive feedback loops, and homeostasis (stability to small perturbations) is based on negative feedback loops.

Regulatory circuits: dynamics in isolation

Positive circuit



Negative circuit





Important relevant readings

Regulatory circuits & Thomas' rules

- A positive regulatory circuit is necessary to generate multiple stable states or attractors
- A negative regulatory circuit is necessary to generate sustained oscillatory behaviour
- Thomas R (1988). *Springer Series in Synergics* **9**: 180-93.

Mathematical theorems and demonstrations:

- In the differential framework:
 - Thomas (+, 1994), Plathe *et al.* (\pm , 1995), Snoussi (\pm , 1998), Gouzé (\pm , 1998), Cinquin & Demongeot (+, 2002), Soulé (+, 2003).
- In the discrete framework:
 - Aracena *et al.* (+, 2001), Remy *et al.* (\pm , 2005), Richard & Comet (+, 2005).

When to use logical modelling?

- There is no quantitative information about the processes (rates of reaction, association/dissociation constants, quantities, etc.)
- The details of some interactions are unknown
- The biological question is qualitative (e.g. how does a cell chooses from a survival or an apoptotic phenotype?)



Discrete models: Boolean networks

Boolean networks have been widely used in modeling gene regulation

Switch-like behaviour of gene regulation resembles logic circuit behaviour

Conceptually easy framework: models easy to interpret

Boolean networks extend naturally to dynamic modeling

Boolean network modeling

- Boolean: either true or false (1 or 0)
- Binarization
 - reduces the noise in biological data
 - captures the dynamic behavior in complex systems
 - need a threshold value
 - leads to loss of information
- Biological entities are modeled as switch like dynamic elements
 - either on or off

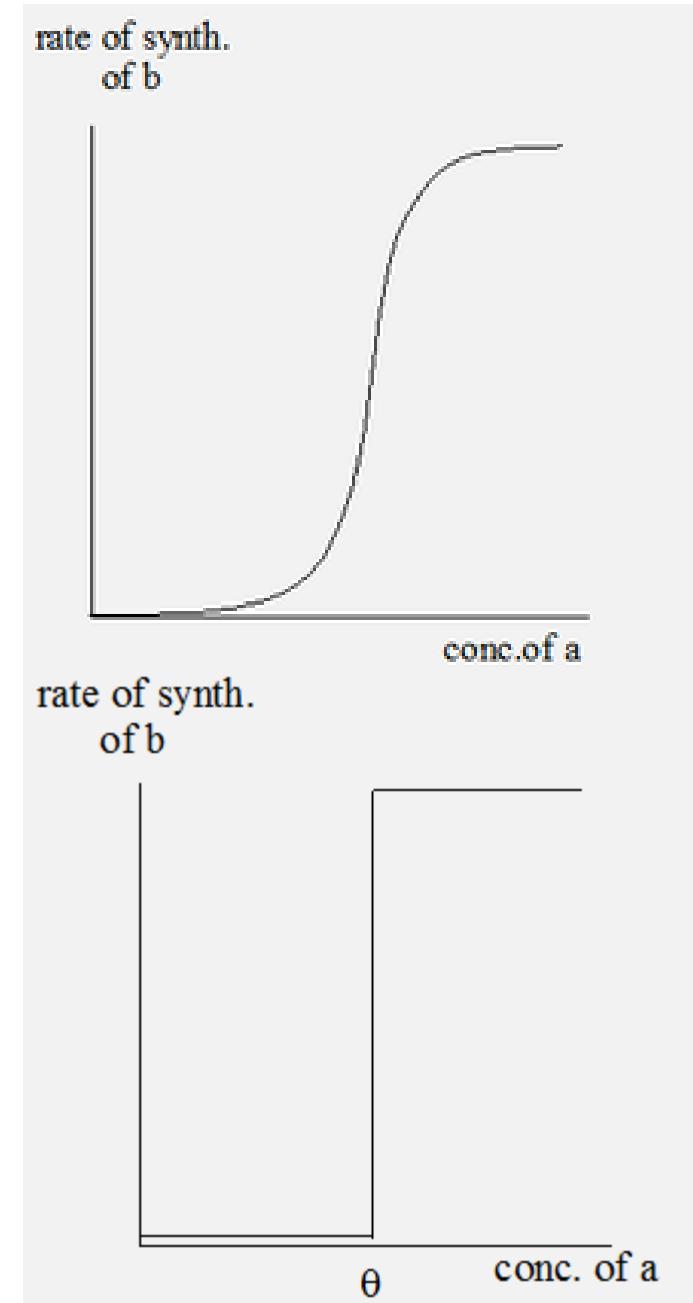
Boolean networks

A Boolean network $G(V, F)$
contains

- Nodes $V = \{x_1, \dots, x_n\}$, $x_i = 0$ or $x_i = 1$
- Boolean functions
 $F = \{f_1, \dots, f_n\}$
- Boolean function f_i is assigned to node x_i
- Dynamic behaviour can be simulated
- State of a variable x_i at time $t+1$ is calculated by function f_i with input variables at time t

Logical Variable and Functions

- Logical variables are associated with the elements of the system to describe the state of the system.
- They consist of the logical values. For example, a system whose state is appropriately described by the levels of substances a, b, and c, each of which can be absent, present at low level, or present at high level are represented by logical values 0, 1, and 2 respectively.
- If a product a acts to stimulate the production of b, it is a positive regulator. In this case, the rate of synthesis of b increases with increasing concentration of a, and makes a curve similar to that shown in figure A.
- There is little effect of a, until it reaches a threshold concentration θ , and at higher concentrations a plateau is reached which shows the maximal rate of synthesis of b.
- Such a nonlinear, bounded curve is called a sigmoid.
- It can be suggested that a is "absent" for $a < \theta$ and "present" for $a > \theta$. The sigmoid curve can be approximated by the step function, as in figure B.



Example of Boolean network dynamics

- Consider a Boolean network with 3 variables x_1 , x_2 and x_3 and functions given by
 - $x_1 := x_2 \text{ and } x_3$
 - $x_2 := \text{not } x_3$
 - $x_3 := x_1 \text{ or } x_2$

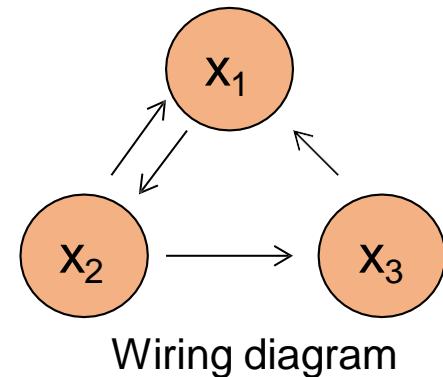
t	x1	x2	x3
0	0	0	0
1	0	1	0
2	0	1	1
3	1	0	1
4	0	0	0
		...	

Example Graph (G) with 3 genes

$G(V, F)$,

$$V = \{x_1, x_2, x_3\}$$

$$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$$



Input (t-1)			Output(t)		
0	0	0			
0	0	1			
0	1	0			
0	1	1			
1	0	0			
1	0	1			
1	1	0			
1	1	1			

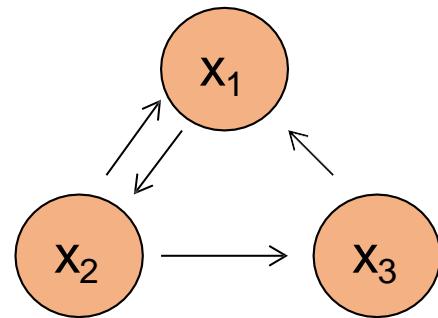
Truth table

Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$



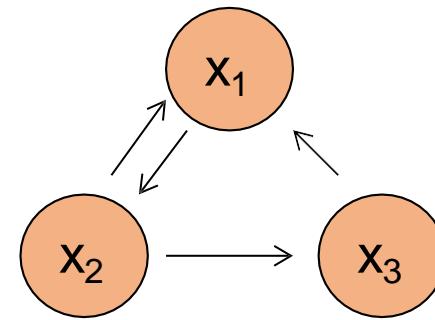
Input (t-1)			Output(t)		
0	0	0	0		
0	0	1			
0	1	0			
0	1	1			
1	0	0			
1	0	1			
1	1	0			
1	1	1			

Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$



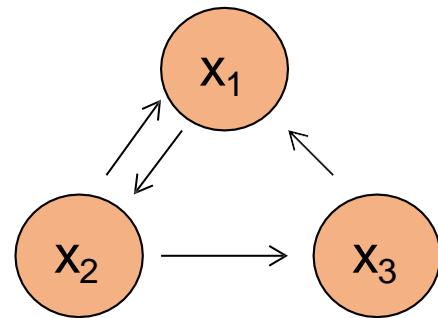
Input (t-1)			Output(t)	
0	0	0	0	0
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	0		
1	1	1		

Example Graph (G) with 3 genes

$G(V, F)$,

$$V = \{x_1, x_2, x_3\}$$

$$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$$



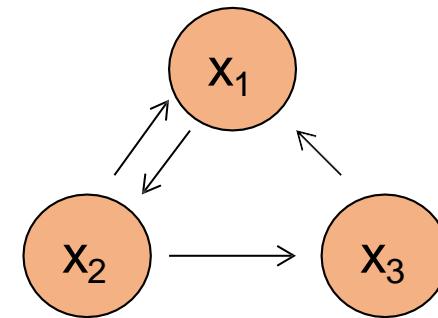
Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1			
0	1	0			
0	1	1			
1	0	0			
1	0	1			
1	1	0			
1	1	1			

Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$



Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1

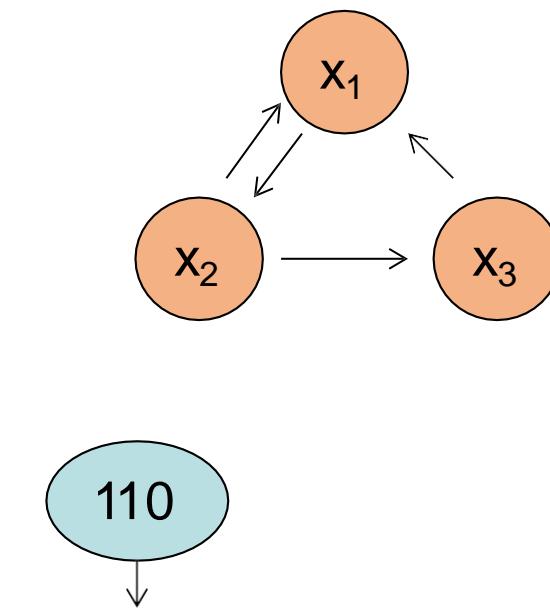
Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$

Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1

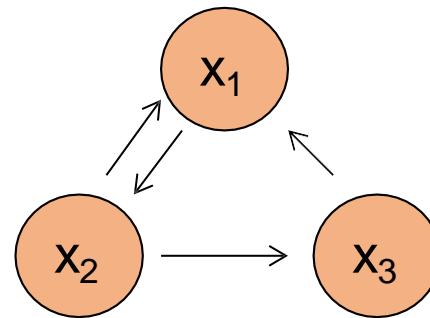


Example Graph (G) with 3 genes

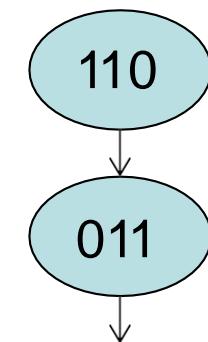
$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$



Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1



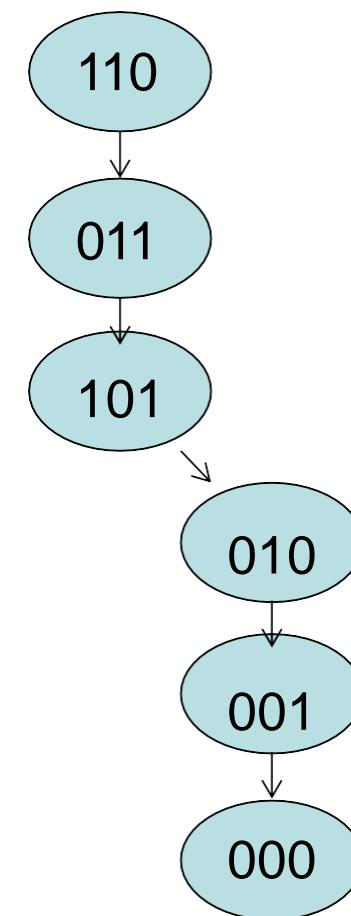
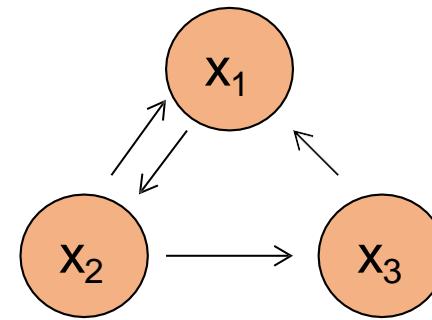
Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$

Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1



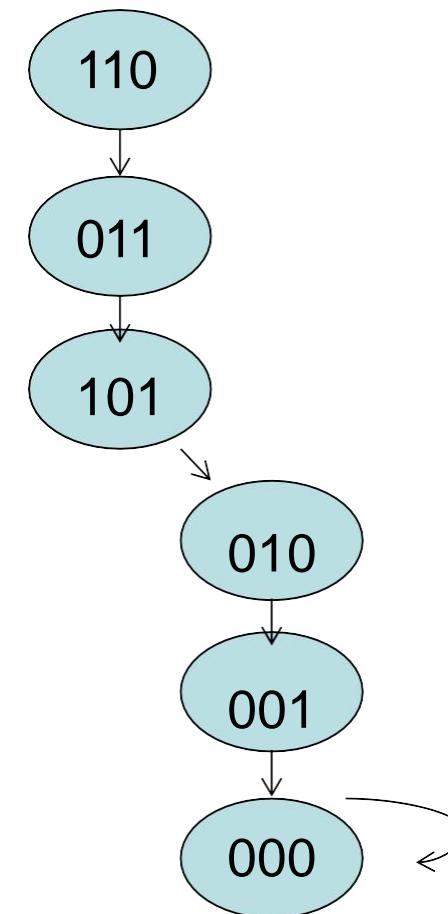
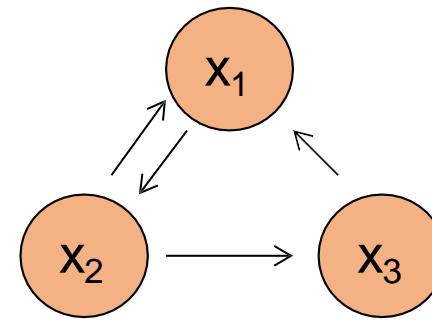
Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$

Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1



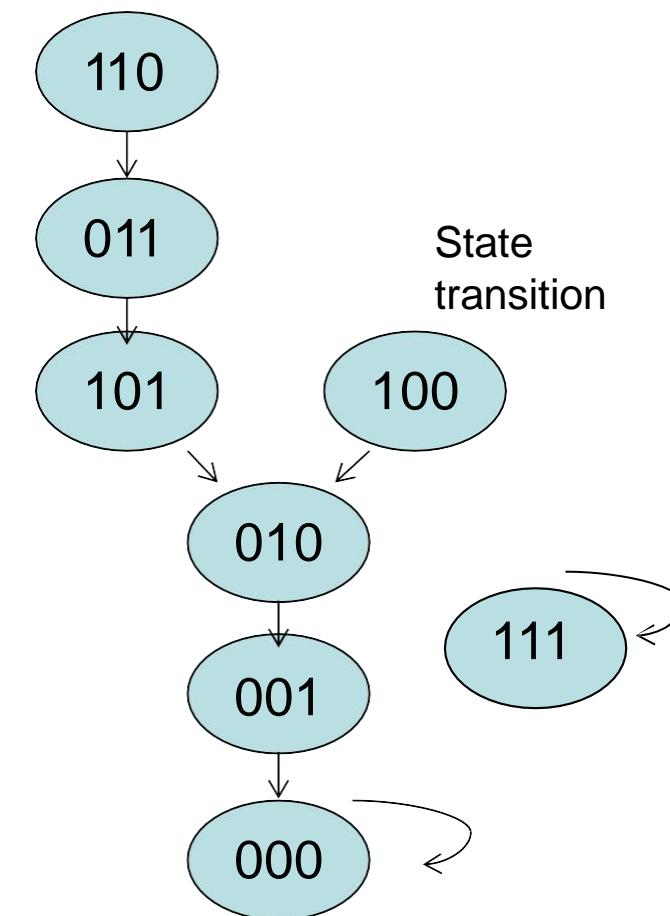
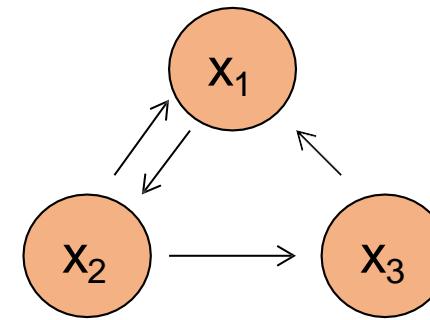
Example Graph (G) with 3 genes

$G(V, F)$,

$V = \{x_1, x_2, x_3\}$

$F = \{f_1 = x_2 \& x_3, f_2 = x_1, f_3 = x_2\}$

Input (t-1)			Output(t)		
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0
1	1	0	0	1	1
1	1	1	1	1	1

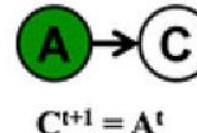


Logic models

- Form of mathematical model governed by logic operators (AND, OR, NOT).
- Parameter free.
- Suitable for modeling gene regulatory networks.
- *In silico* simulations, qualitative predictions.
- Each node in a logic model has a corresponding logic function that controls its regulation each time the model is updated.
- Two updating schemes: synchronous and asynchronous.

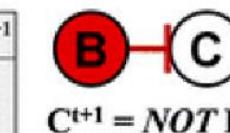
A Logic functions with one molecular regulator

A^t	C^{t+1}
0	0
1	1



$$C^{t+1} = A^t$$

B^t	C^{t+1}
0	1
1	0



$$C^{t+1} = \text{NOT } B^t$$

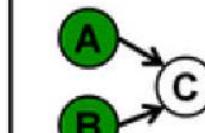
Truth table

Truth table

Logic functions with two molecular regulators

B Non-specific Interaction Network

Two Activators



$$C^{t+1} = A^t \text{ AND } B^t$$

A^t	B^t	C^{t+1}
0	0	0
1	0	0
0	1	0
1	1	1

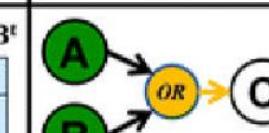
The presence of A and the presence of B activates C

AND

C is only ON in one condition

OR

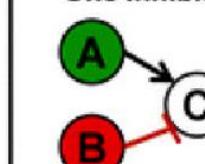
C is only OFF in one condition



A^t	B^t	C^{t+1}
0	0	0
1	0	1
0	1	1
1	1	1

Either the presence A or the presence of B activates C.

One Activator and One Inhibitor

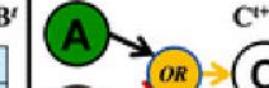


$$C^{t+1} = A^t \text{ AND NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	0
1	0	1
0	1	0
1	1	0

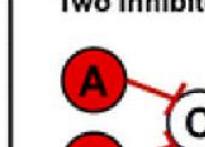
The presence of A and the absence of B activates C.
Inhibitor Dominant

A^t	B^t	C^{t+1}
0	0	1
1	0	1
0	1	0
1	1	1



Either the presence of A or the absence of B activates C.
Activator Dominant

Two Inhibitors



$$C^{t+1} = \text{NOT } A^t \text{ AND NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	1
1	0	0
0	1	0
1	1	0

The absence of A and the absence of B activates C.

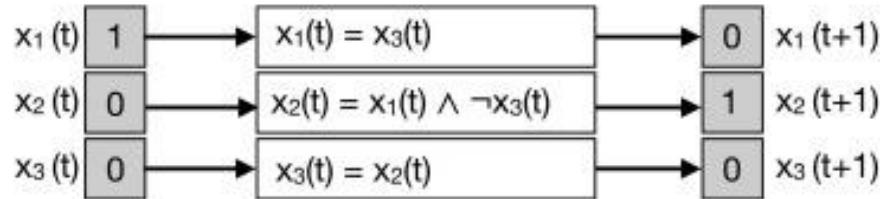
A^t	B^t	C^{t+1}
0	0	1
1	0	1
0	1	1
1	1	0



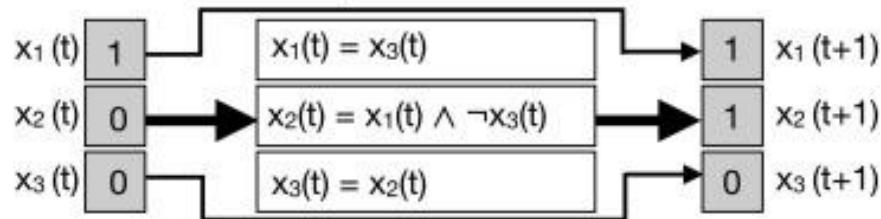
Either the absence of A or the absence of B activates C.

Updating schemes

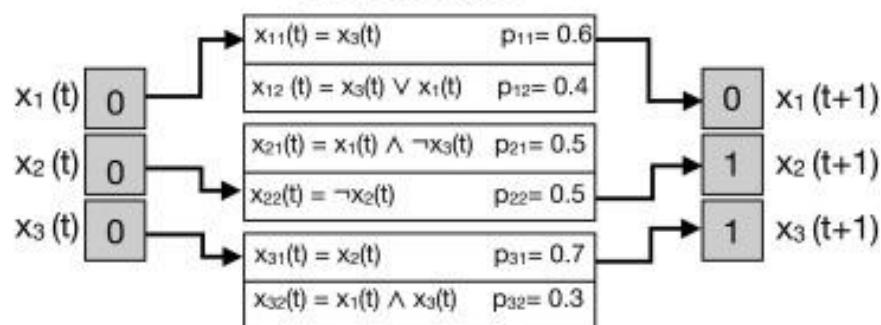
synchronous



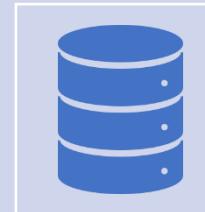
asynchronous



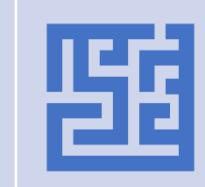
probabilistic



In synchronous models all Boolean functions are applied at the same time while in asynchronous models only one randomly chosen function (fat arrow) is updated per step.



Probabilistic models can have multiple functions with a predefined probability. One function per variable x_i is randomly chosen in each time step and then synchronous updating is performed



stable state (terminal nodes in the state transition graphs)

cyclic attractor (terminal strongly connected components)

transient cycle(s) (non-terminal strongly connected components)

basins of attraction (subsets of transient states)

1. Biological
Regulatory
Network

2. Logical
Variables &
Functions

3. Graph of
Interactions and
Logical
Equations

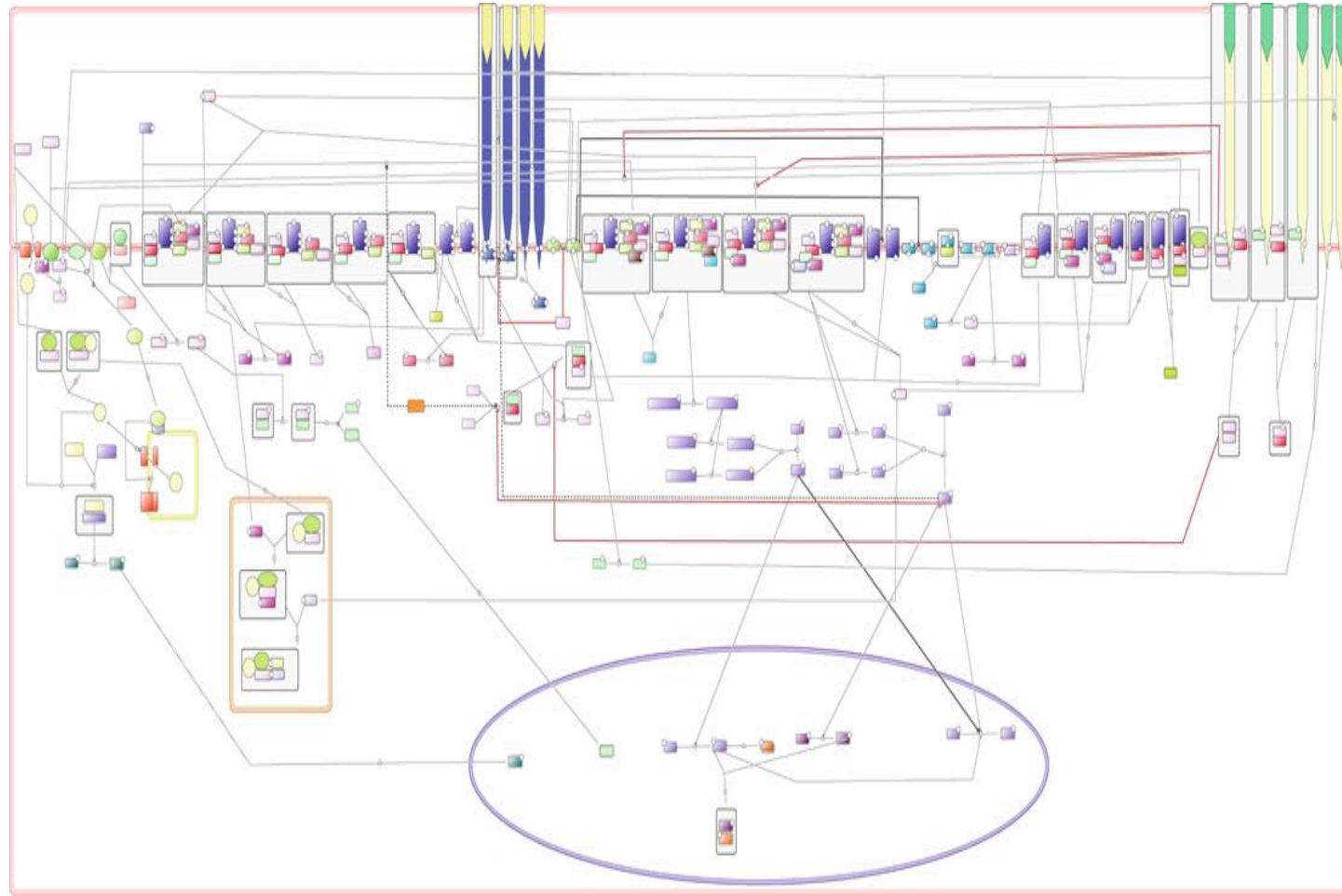
6. Analysis

5. Cycles and
Stable Steady
States

4. State Table
and State Graph

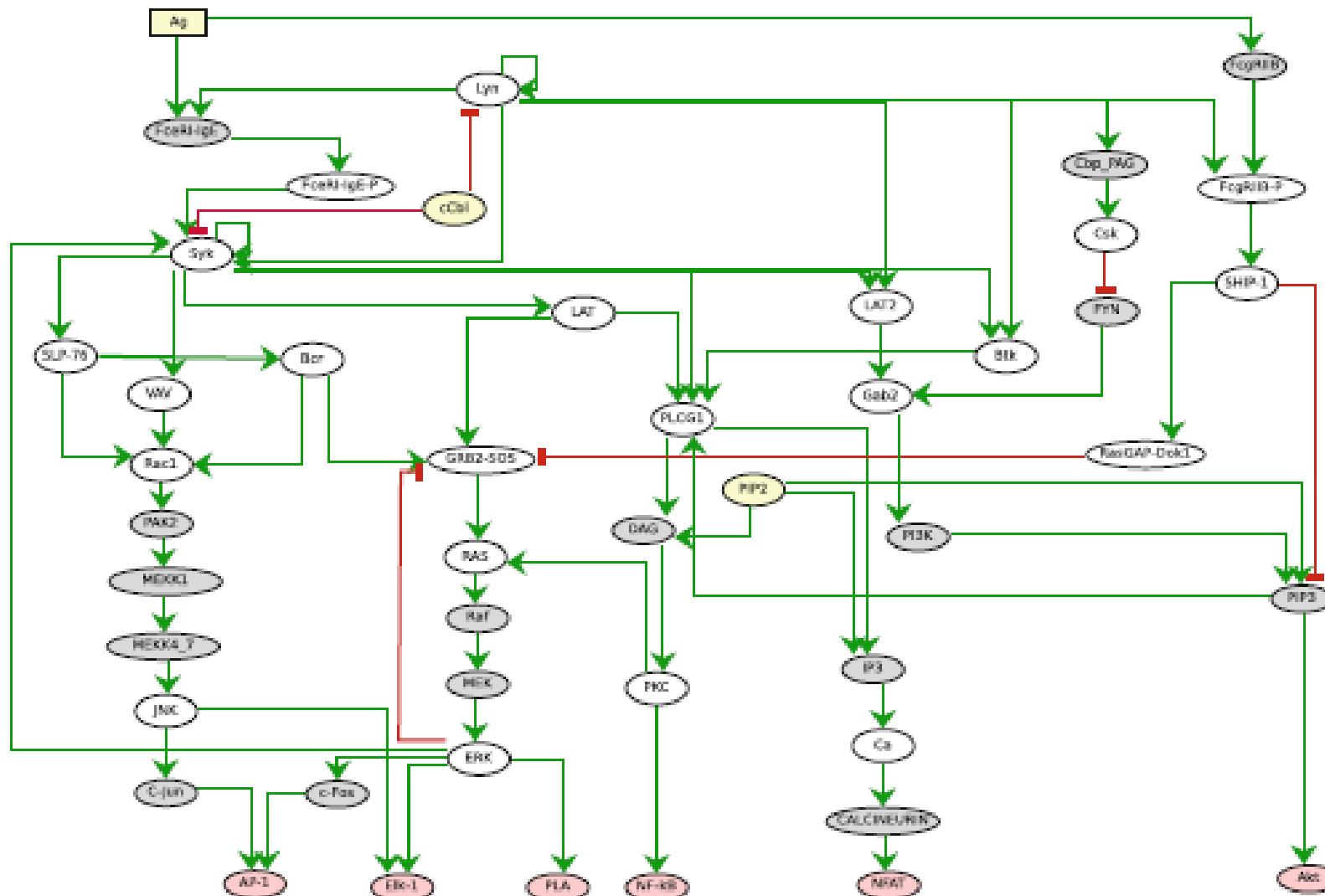


Mast cell activation signaling pathway map created with CellDesigner



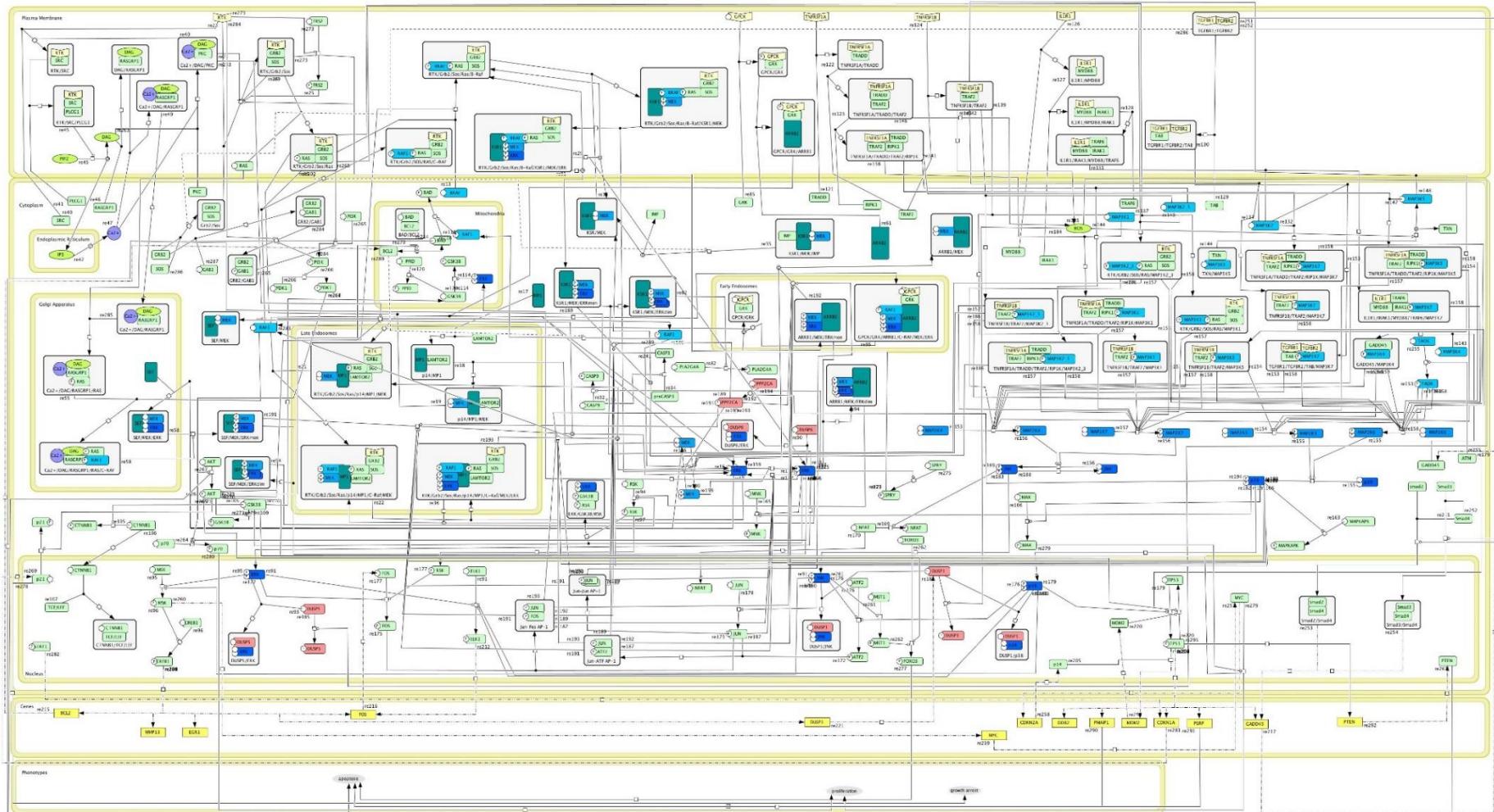
Niarakis, Anna, et al. "Computational modeling of the main signaling pathways involved in mast cell activation." *Fc Receptors*. Springer International Publishing, 2014. 69-93.

Regulatory graph of the mast cell signaling logical model



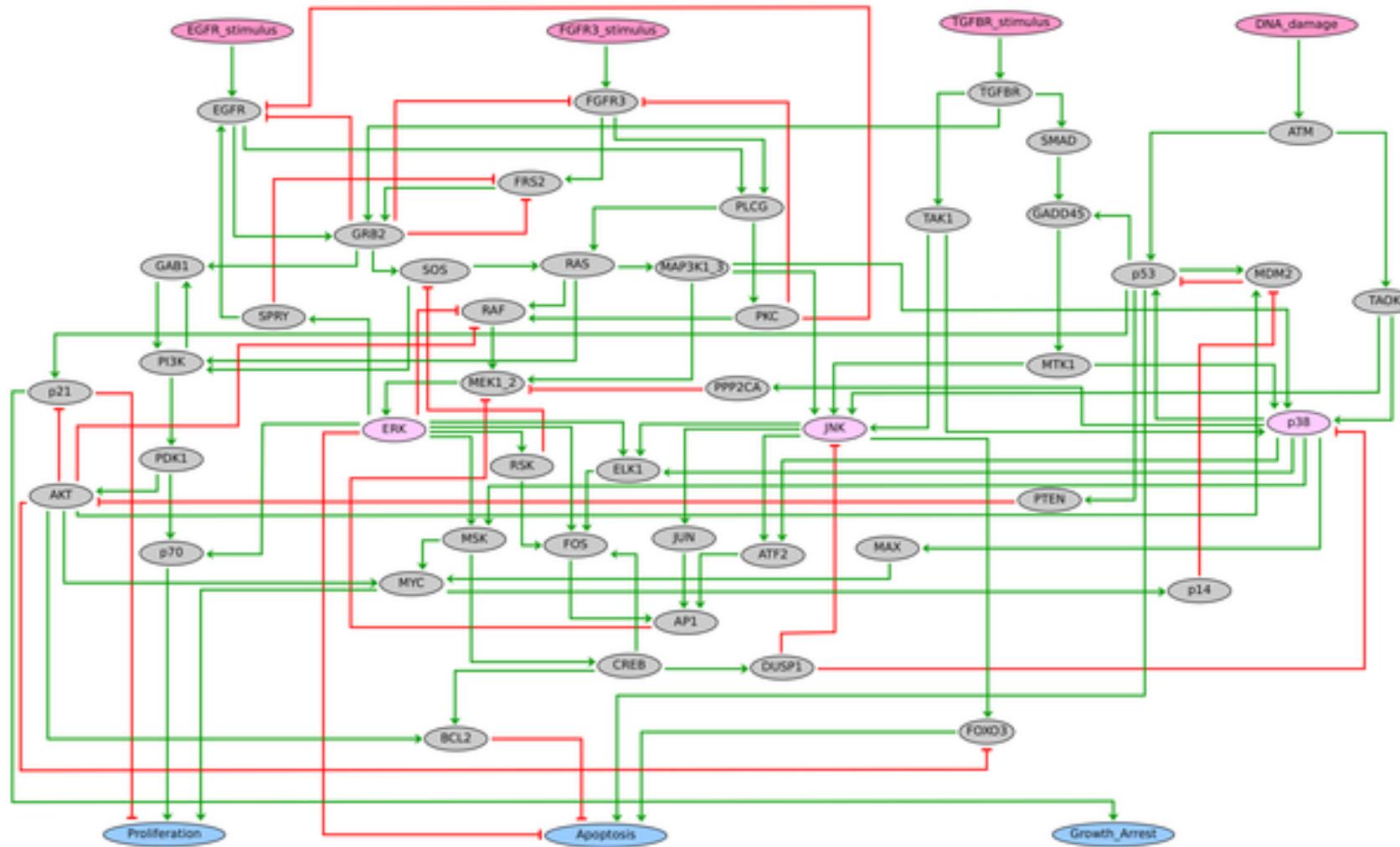
Niarakis, Anna, et al. "Computational modeling of the main signaling pathways involved in mast cell activation." *Fc Receptors*. Springer International Publishing, 2014. 69-93.

Molecular map for MAPK pathway



Grieco L, Calzone L, Bernard-Pierrot I, Radvanyi F, Kahn-Perlès B, et al. (2013) Integrative Modelling of the Influence of MAPK Network on Cancer Cell Fate Decision. PLoS Comput Biol 9(10): e1003286. doi:10.1371/journal.pcbi.1003286
<http://journals.plos.org/ploscompbiol/article?id=info:doi/10.1371/journal.pcbi.1003286>

Regulatory graph of the MAPK logical model.

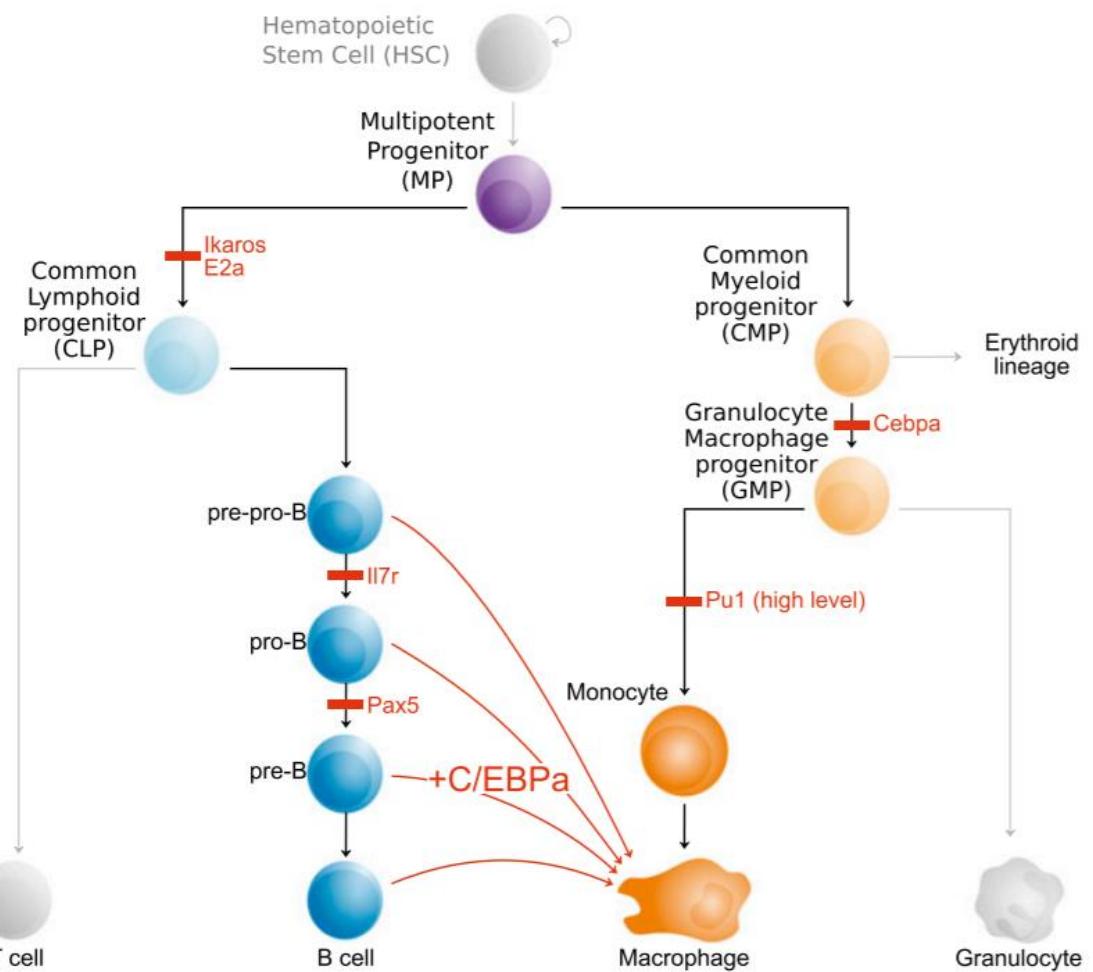


Grieco L, Calzone L, Bernard-Pierrot I, Radvanyi F, Kahn-Perlès B, et al. (2013) Integrative Modelling of the Influence of MAPK Network on Cancer Cell Fate Decision. PLoS Comput Biol 9(10): e1003286. doi:10.1371/journal.pcbi.1003286
<http://journals.plos.org/ploscompbiol/article?id=info:doi/10.1371/journal.pcbi.1003286>

Logical modeling of lymphoid and myeloid cell specification and transdifferentiation

Samuel Collombet^{a,1}, Chris van Oevelen^{b,2}, Jose Luis Sardina Ortega^{b,2}, Wassim Abou-Jaoudé^a, Bruno Di Stefano^{b,3}, Morgane Thomas-Chollier^a, Thomas Graf^{b,c,1}, and Denis Thieffry^{a,1}

A



B

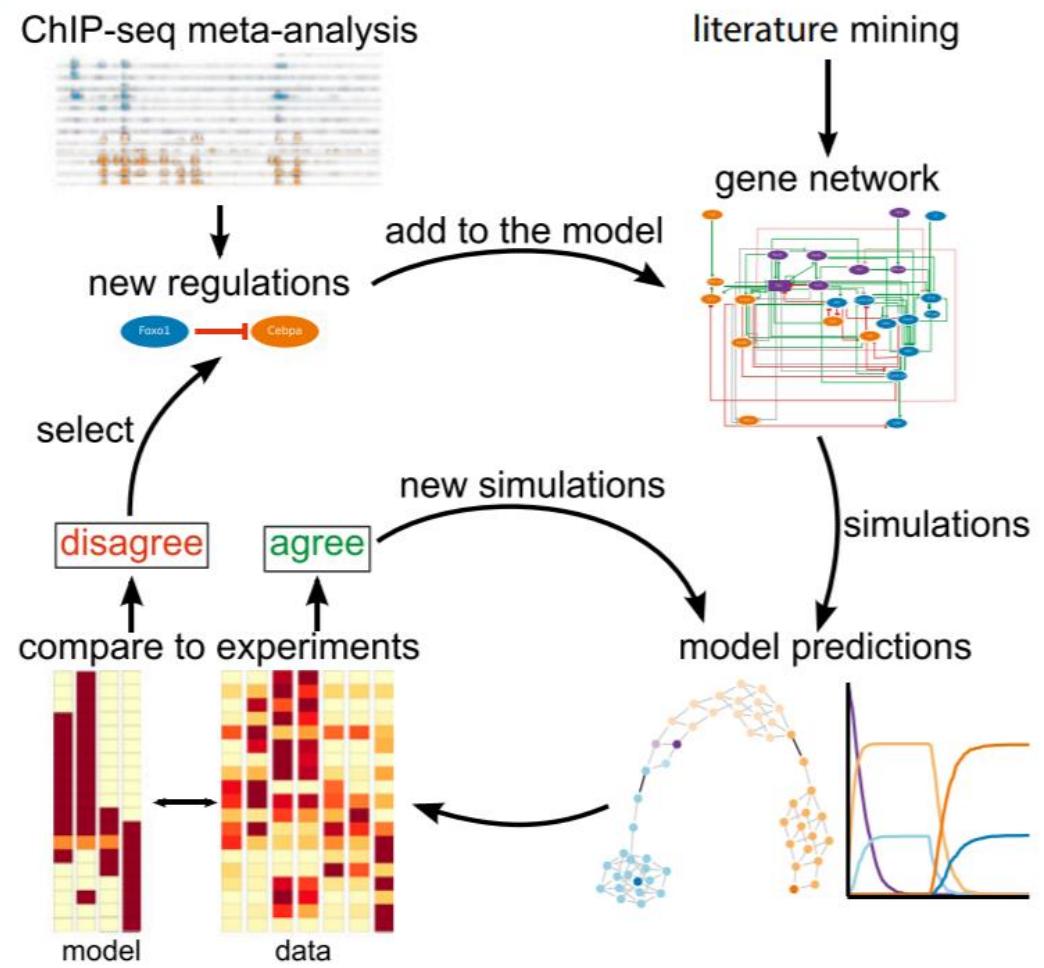
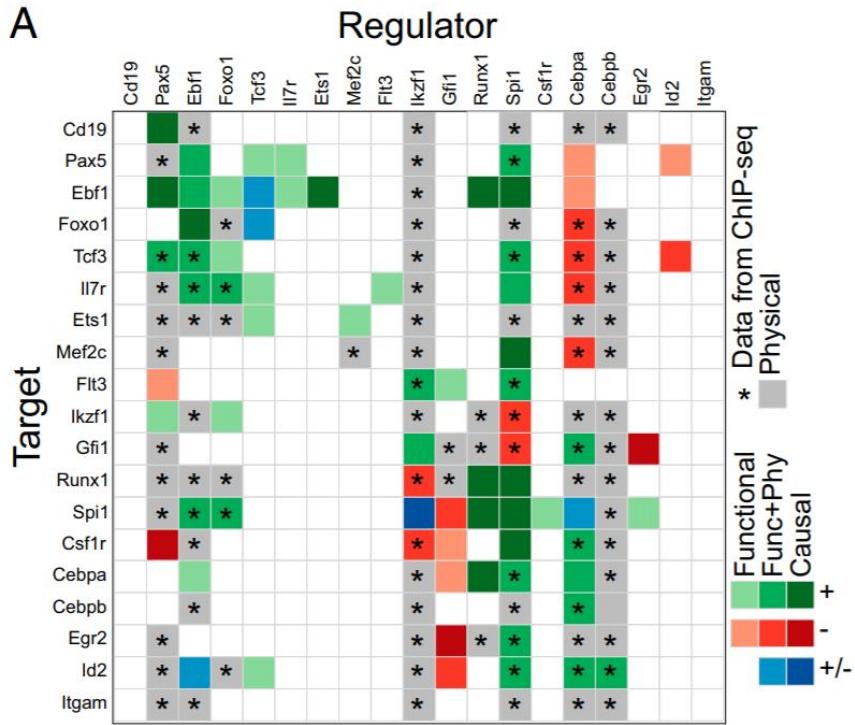
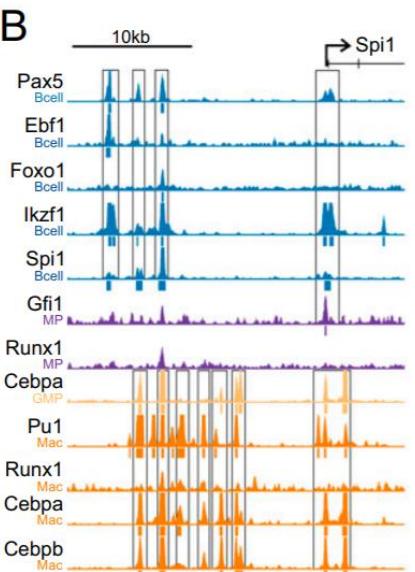


Fig. 1. (A) Schematic representation of hematopoietic cell specification. Genes in red are required for progression at the corresponding steps. C/EBP α -induced transdifferentiation is indicated by red arrows from B-lineage cells to macrophages. (B) Iterative modeling workflow. A model is first built based on the literature and is used to predict dynamical behaviors (cell phenotype, differentiation, reprogramming, and so forth). Predictions then are compared with experimental data; when the predictions and experimental data agree, further predictive simulations are performed; when they do not agree, further regulations are inferred from ChIP-seq data and are integrated into the model until simulations fully agree with data.

A



B



C

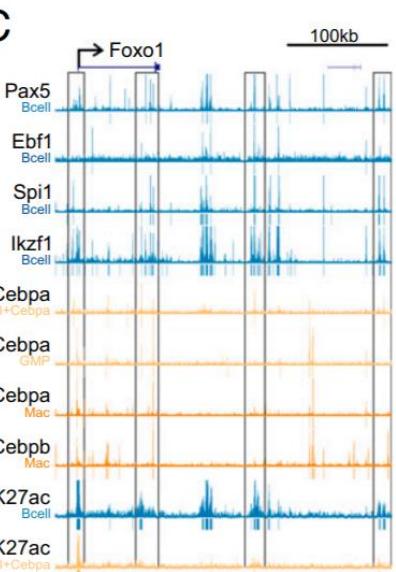


Fig. 2. (A) Heatmap showing the regulations inferred from the literature and from ChIP-seq meta-analysis. (B) ChIP-seq signals and peaks (under signal) at the *Spi1* locus. Black frames indicate known enhancers (24). The vertical axes represent reads per million (RPM) (maximum: 2 RPM for Ebf1 and Ikaros, 1.5 RPM for Foxo1, 1 RPM for Runx1 and Gfi1, 5 RPM for other TF). (C) ChIP-seq signals and peaks (under signal) at the *Foxo1* locus. Black frames indicate B-cell enhancers in which C/EBP α binding is detected. The vertical axes represent RPM (maximum: 2 RPM for Ebf1, 5 RPM for other TFs, 3 RPM for H3K27ac). Note that Pax5 and Ikaros peaks are located downstream of the first exon and all other peaks are upstream of the TSS.

A regulatory graph depicting the interactions inferred from the literature and ChIP-seq meta-analyses.

Nodes represent genes (except for CSF1r_act and IL7r_act, which represent the activated forms of cytokine receptors), and arrows denote regulatory interactions.

Orange nodes represent factors expressed in macrophages, purple nodes represent factors expressed in progenitors, and blue nodes represent factors expressed in B-lineage cells.

Ellipses represent Boolean components; the rectangle emphasizes the ternary component Spi1.

Green and red edges correspond to activations and inhibitions, respectively.

Gray edges denote the regulations predicted by the ChIP-seq meta-analysis, which were included in the model to increase consistency with expression data

